Published online 18 November 2011 *Nucleic Acids Research, 2012, Vol. 40, Database issue* ***D525–D530***
doi:10.1093/nar/gkr1021

STAP Refinement of the NMR database: a database of 2405 refined solution NMR structures

Joshua SungWoo Yang[1,2], Ji-han Kim[1], Sangho Oh[1], Gukjeong Han[1], Sanghyuk Lee[1,3] and Jinhyuk Lee[1,2,*]

[1]Korean Bioinformation Center (KOBIC), Korea Research Institute of Bioscience and Biotechnology, 125 Gwahak-ro Yuseong-Gu, Daejeon 305-806, [2]Department of Bioinformatics, University of Science and Technology, 217 Gajung-ro Yuseong-Gu, Daejeon 305-350 and [3]Ewha Research Center for System Biology (ERCSB), Ewha Woman's University, 11-1 Daehyun-dong, Seodaemun-gu, Seoul 120-750, The Republic of Korea

Received August 15, 2011; Revised October 20, 2011; Accepted October 21, 2011

ABSTRACT

According to several studies, some nuclear magnetic resonance (NMR) structures are of lower quality, less reliable and less suitable for structural analysis than high-resolution X-ray crystallographic structures. We present a public database of 2405 refined NMR solution structures [statistical torsion angle potentials (STAP) refinement of the NMR database, http://psb.kobic.re.kr/STAP/refinement] from the Protein Data Bank (PDB). A simulated annealing protocol was employed to obtain refined structures with target potentials, including the newly developed STAP. The refined database was extensively analysed using various quality indicators from several assessment programs to determine the nuclear Overhauser effect (NOE) completeness, Ramachandran appearance, χ_1-χ_2 rotamer normality, various parameters for protein stability and other indicators. Most quality indicators are improved in our protocol mainly due to the inclusion of the newly developed knowledge-based potentials. This database can be used by the NMR structure community for further development of research and validation tools, structure-related studies and modelling in many fields of research.

INTRODUCTION

Protein structure determination has contributed greatly to our understanding of structural biology by providing the three-dimensional protein structure and the role of the associated conformational dynamics of the protein. Despite the importance of protein structures, three-dimensional protein structure determination has been limited to X-ray crystallography in the solid state (single crystals) or nuclear magnetic resonance (NMR) spectroscopy in solution. Recently, 9049 macromolecular structures obtained by NMR spectroscopy were available in the Protein Data Bank (PDB) (1,2). NMR is uniquely suited to the characterization of protein structure and dynamics in solution, and it is not hampered by the ability or inability of a protein to crystallize. Despite these advantages, NMR protein structures are usually not a first choice for studies of protein structure and function. Because NMR structures can be of lower quality, less reliable and less suitable for structural analysis than crystallography, they are often excluded (3–5).

Because of the uncertainty in NMR structures, numerous refinement protocols and force fields have been developed to improve the quality of NMR structures (3,5–7). Even with the recent advancements in protocols and force fields, the structures obtained can still be of unsatisfactory quality, indicating that there is room for further improvement. Numerous studies have advised that the poor quality of the Ramachandran plot, backbone conformation and/or side-chain packing were caused by the low quality of the NMR structures (8–10). To address these weaknesses, re-refinement and other protocols have been introduced in several projects [DRESS (11), RECOORD (12), etc.]. Re-refinement usually leads to an improvement in structure quality.

The re-refinement in this study is focused on 2405 of the 9049 selected solution NMR structures deposited in the PDB. We developed a new refinement protocol to refine the solution NMR structure by correcting the backbone polypeptide torsion angles. A detailed protocol and target selection are introduced in the next section. We illustrate the clear overall improvement of the structures in our database, statistical torsion angle potentials (STAP) refinement of the NMR database, compared to the currently available refinement databases.

*To whom correspondence should be addressed. Tel: +82 42 879 8530; Fax: +82 42 879 8519; Email: jinhyuk@kribb.re.kr

STAP refinement of NMR database

Unlike crystallography, NMR structure determination uses very heterogeneous geometrical information: distances, dihedral angles and orientations (12). Among this geometrical information, we focused on the dihedral angle. Torsion angles are assumed to be a very important factor in the quality of NMR structures (8–10). Its deviation sways the inter-atomic distance and greatly influences the nuclear Overhauser effect (NOE) signals in experiments (13,14). Torsion angle population analysis provides a firm ground for deriving knowledge-based energy functions. STAP was developed with 18 353 high-resolution X-ray crystallographic structures with a resolution below 2.0 Å from the PDB. From that, we removed the structures with redundancies and built two-dimensional histograms as a function of the two backbone torsion angles (φ and ψ) with a grid point for every 15°. By applying a log transformation to these histograms (15), we obtained the two-dimensional knowledge-based potentials on φ/ψ torsion angles for favourable conformational isomerism. We call this potential STAP. STAP gives weight to suitable conformations and lets preferable ones remain. It especially works with the simulated annealing protocol.

The 2405 structures are presented in the STAP refinement of the NMR database. Most of the experimental structures available from the PDB were loaded directly into the automatic pipeline for refinement. Each NOE distance restraint was downloaded from the Biological Magnetic Resonance Bank (BMRB) (16). This restraint set was used to refine the corresponding ensembles deposited in the PDB. Targets that carry contradicting inter-atomic distances to the NOE data were discarded. We selected only the X-PLOR restraint format.

Briefly, the refinement protocol used is as follows: (i) after STAP refining potentials are applied to all of the structures, the implicit solvation model is applied (17) and energy-minimized; (ii) the system is heated from 100 to 500 K using 1000 steps of molecular dynamics; (iii) 2000 steps of molecular dynamics at 500 K are performed; (iv) cooling down to 25 K runs during 4000 steps; and (v) a short minimization with 200 steps is performed. This protocol was executed using CHARMM (ver. C35) (18).

Quality assessment of NMR structure

The programs PROCHECK (ver. 3.5.4) (8) and MolProbity (ver. 2.12) (10) were used to measure the Ramachandran appearance and steric clash score. The program WHAT_CHECK (ver. 8.0) (9) was used to measure the root mean square (RMS) *Z*-score distribution of several parameters of the protein structure. There are three optimal energy properties for protein structures: Discrete Optimised Protein Energy (DOPE, ver. 9v7) (19), normalized DOPE (nDOPE, ver. 9v7) (20) and dipolar Distance-scaled, Finite-Ideal gas REference (dDFIRE, ver. 1.1) (21). NOE distance violations for all ensemble restraints were calculated with AQUA (ver. 3.2) (22).

Statistical analysis for STAP refinement of the NMR database

Statistical RMS *Z*-score distribution, RMS NOE violations and nDOPE scores for the original and refined structure are shown in Figure 1. As we expected, the RMS *Z*-score for the Ramachandran plot appearance (Figure 1D) showed a great improvement because it shows two separate Gaussian distributions. In addition, the result of the RMS *Z*-score distribution for the χ_1–χ_2 rotamer normality (Figure 1E) also appears similar to that of the Ramachandran plot appearance. Other parameters show remarkable improvement with a stable low energy profile for nDOPE and 2nd-generation packing quality. Interestingly, the influence of the torsion angle conformation on the NMR structure may cause RMS NOE violations to provide better quality, indicating that NOE violations directly affect the geometrical conformation of the torsion angle (23,24).

The average values of the original and refined parameters pertaining to the structural quality are presented in Table 1. Evaluation of the structural quality indicated that our protocol, STAP, provides improved quality in the polypeptide backbone conformation. Remarkably, the number of RMS NOE violations was closer to zero in the refined structure than in the original structure, which is shown above in the histogram depicted in Figure 1.

Comparison with DRESS and RECOORD

From analysis of the STAP data, there are 70 structures present in the two public NMR refined databases (DRESS and RECOORD databases), as shown in Table 2. The refined data of DRESS (refined PDB files) and RECOORD (CNW and CYW PDB files) were loaded directly from their websites to measure the protein structure qualities. The quality improvement by the three protocols (STAP, DRESS and RECOORD) is shown in Table 2. The STAP refinement protocol has a great improvement on the Ramachandran backbone conformation, especially on the percentage of favourable indicators by MolProbity and PROCHECK. The *Z*-score distributions by WHAT_CHECK indicated that our protocol provides slightly better quality than the other two protocols [steric clash score, nDOPE score, RMS *Z*-score distributions on the structure (2nd-generation packing quality, χ_1–χ_2 rotamer normality, Ramachandran plot appearance and backbone conformation) and RMS *Z*-score distributions (bond angles and inside/outside distribution)]. The result of the optimized energy comparison indicated that the other databases (RECOORD and DRESS) are slightly better than our approach. With the many positive attributes of STAP, it can be confidently concluded that STAP provides comparable performance to justify the geometric consensus for torsion angles and shows comparable or slightly better performance than the two known protocols.

We have discussed the statistical analysis of the original and refined structures and compared our refined structure quality with that of other known NMR refined databases. Based on these facts, we believe that the quality of our 2405 NMR structure database is significantly improved,

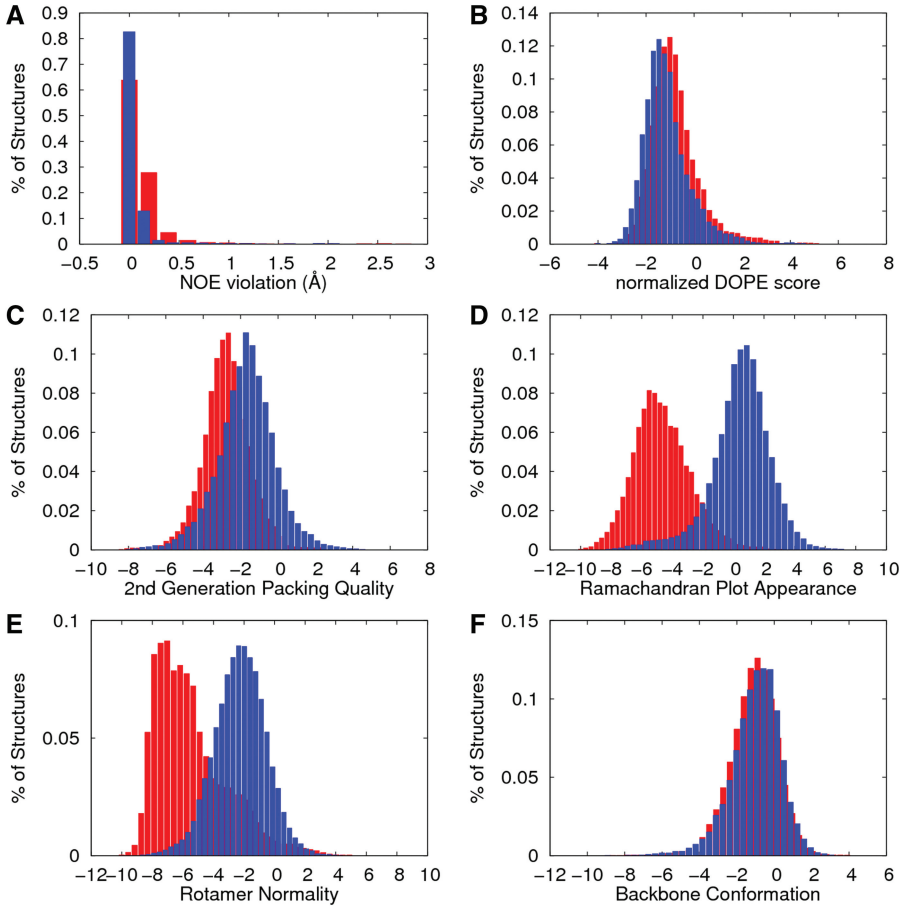


Figure 1. Distribution of the protein quality assessment of the original (red) and refined structures (blue bars). The figures present the distributions of the 2405 selected NMR structures as follows: (**A**) the RMS value of the NOE violations, (**B**) the normalized DOPE score, (**C**) the 2nd-generation packing quality (all backbone and side chain contacts), (**D**) the Ramachandran plot appearances, (**E**) the χ_1–χ_2 rotamer normality and (**F**) the backbone conformation. The values were measured by dDFIRE, AQUA and WHAT_CHECK.

and this database will be a good start for further development of research and validation tools, structure-related studies and modelling in many fields of research.

Web interface and system

The STAP refinement of the NMR database can be accessed using a web-portal (http://psb.kobic.re.kr/STAP/refinement) that provides a 2405 refined NMR structure database, which runs on the CentOS operation system (ver. 5.2). The web server uses a standard web browser as the graphic user interface to the front-end server based on PHP (ver. 5.1.6)/SQL, the application server based on Apache (ver. 2.2.3) and the MySQL (ver. 5.0.45) back-end database server.

The following five website features are described in detail: (i) the main page, (ii) the general structure information page, (iii) the results page, (iv) the visualization page and (v) the download page. (i) The Front page allows the user to enter the selected PDB ID to access the database. On the main page, there are three different ways to access the database (PDB List, Lucky and manual). The PDB List is somewhat similar to the MS Windows file explorer interface, which is one of the search engines using the tree structure as the user interface, sequentially listing PDB IDs. If a user clicks the 'PDB List' button, a tree browser layer will open at the upper left side of the window on the main page. The top-level directory consists of the first letter of each PDB ID. When the user clicks the '+' button placed on the left side of the directory image, the next lower level of the tree node appears. When the user clicks 'first two letters of ID', the user can see the list of PDB IDs. The second way is using 'Lucky', which is known as the 'online demonstration', and uses the pseudo-random number generator algorithms (25) to select a random target. 'Lucky' provides visual information with a mouse–rollover–tooltip layer, so that a user can learn our system quickly. Finally, for the manual search, a user types the PDB ID to access the database. For user convenience, the auto-completion function is available. (ii) General information on the target NMR structure (Figure 2) is provided, including information on the protein, experimental information, visualization of the structure using the Jmol viewer (ver. 11.4.RC4) (26) and links to the PDB, BMRB, PubMed (dependent on PDB data; if the PDB data do not include the PubMed information then the PubMed icon

Table 1. Summary of the average quality indicators of the original and refined NMR structures

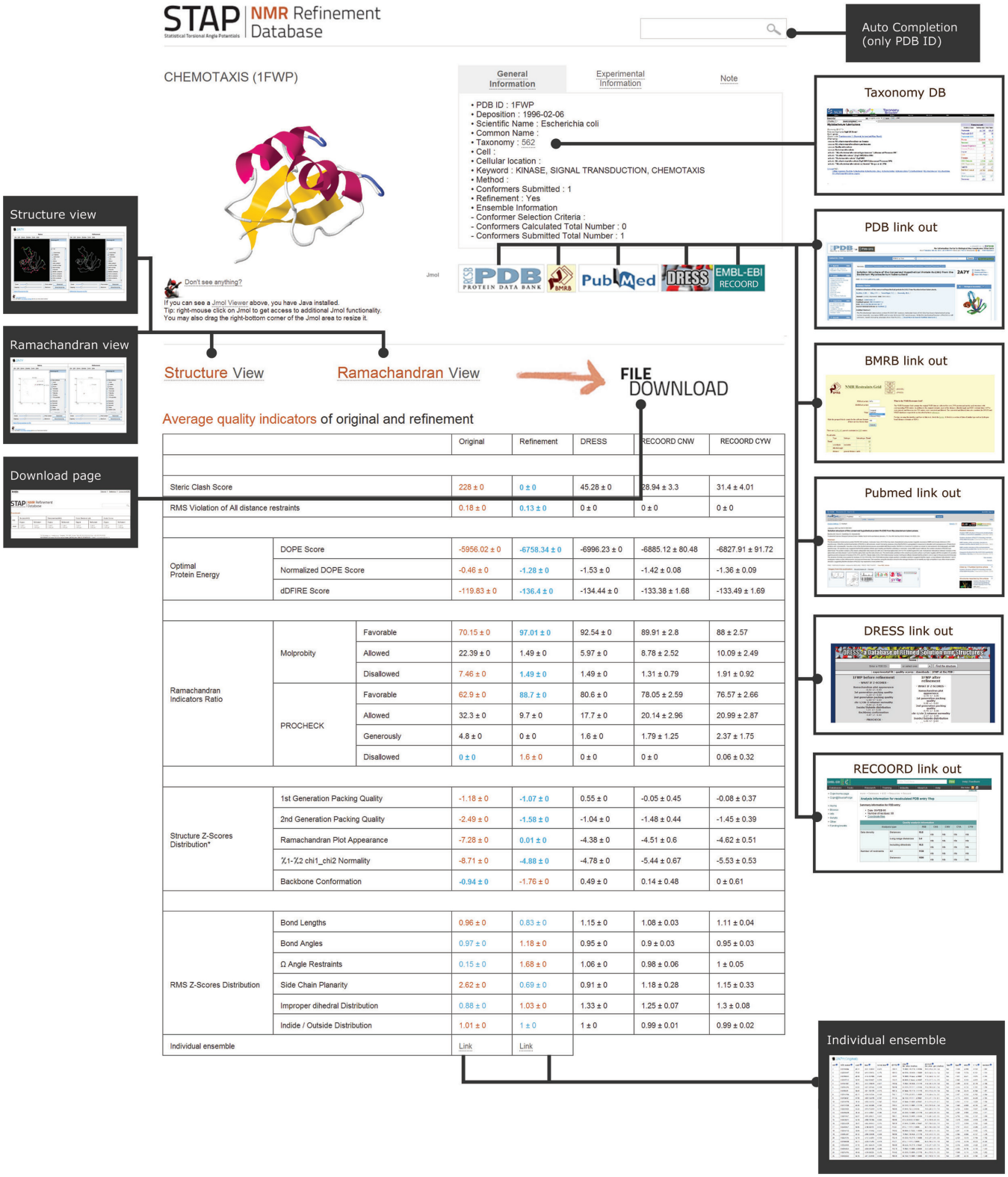
	Original	Refined (STAP)[a]
Steric clash score	33.16 ± 39.52	**1.57 ± 5.90**
RMS NOE distance violations	0.20 ± 0.35	**0.13 ± 0.24**
Optimal protein energy		
DOPE score	−8050.74 ± 4572.84	**−8587.34 ± 4817.02**
Normalized DOPE score	−0.62 ± 1.02	**−1.02 ± 0.97**
dDFIRE Score	−163.12 ± 89.44	**−179.07 ± 96.77**
Ramachandran indicators (%)		
MolProbity		
Favourable	83.50 ± 10.09	**95.51 ± 4.83**
Allowed	12.95 ± 7.36	3.80 ± 3.92
Disallowed	3.54 ± 4.04	0.69 ± 1.48
PROCHECK		
Favourable	75.66 ± 12.08	**89.83 ± 7.45**
Allowed	21.04 ± 9.65	8.01 ± 5.80
Generously	2.35 ± 3.31	0.84 ± 1.55
Disallowed	0.96 ± 1.72	1.32 ± 2.01
Structure *Z*-score distribution[b]		
1st-generation packing quality	−3.14 ± 1.68	**−2.63 ± 1.80**
2nd-generation packing quality	−2.59 ± 1.27	**−1.52 ± 1.55**
Ramachandran plot appearance	−4.50 ± 1.86	**0.65 ± 1.95**
χ_1–χ_2 Rotamer normality	−5.25 ± 2.36	**−2.07 ± 1.68**
Backbone conformation	−0.93 ± 1.28	**−0.80 ± 1.33**
RMS *Z*-score distribution[c]		
Bond lengths	0.48 ± 0.34	**0.84 ± 0.12**
Bond angles	0.70 ± 0.39	**1.04 ± 0.16**
Ω Angle restraints	0.36 ± 0.51	**1.30 ± 0.43**
Side-chain planarity	0.59 ± 0.92	**0.68 ± 0.45**
Improper dihedral distribution	0.64 ± 0.39	**0.79 ± 0.25**
Inside/outside distribution	1.05 ± 0.10	**1.03 ± 0.10**

[a]Bold font indicates the best scores.
[b]Positive is better than average.
[c]RMS *Z*-score should be close to 1.0.

Table 2. Comparison of the 70 NMR structures that are common among the STAP, DRESS and RECOORD databases

	Original	STAP[a]	DRESS	RECOORD-CNW	RECOORD-CYW
Steric clash score	68.26 ± 57.86	**1.03 ± 2.44**	16.79 ± 9.56	16.31 ± 8.62	16.56 ± 8.46
Optimal protein energy					
DOPE score	−6098.15 ± 4586.33	−6941.80 ± 5130.94	−6996.31 ± 5244.70	−7274.30 ± 5902.43	**−7330.00 ± 6150.59**
Normalized DOPE score	−0.21 ± 1.17	−1.00 ± 1.06	**−1.01 ± 1.13**	−0.95 ± 1.11	−0.96 ± 1.10
dDFIRE score	−125.95 ± 88.72	−147.52 ± 102.79	−143.07 ± 101.01	−146.95 ± 114.19	**−148.66 ± 120.39**
Ramachandran indicators (%)					
MolProbity					
Favourable	75.02 ± 12.37	**94.80 ± 4.51**	86.38 ± 8.18	84.69 ± 9.00	83.59 ± 9.35
Allowed	18.76 ± 8.87	4.27 ± 3.64	10.71 ± 6.71	11.89 ± 7.03	12.75 ± 7.58
Disallowed	6.22 ± 5.25	0.92 ± 1.55	2.90 ± 2.92	3.43 ± 3.34	3.66 ± 3.29
PROCHECK					
Favourable	67.02 ± 15.14	**88.45 ± 7.60**	77.92 ± 11.70	75.33 ± 13.02	73.73 ± 13.88
Allowed	27.18 ± 11.95	8.73 ± 6.08	18.52 ± 9.79	20.20 ± 10.73	21.48 ± 11.54
Generously	4.38 ± 4.10	1.00 ± 1.59	2.06 ± 2.49	2.70 ± 3.00	2.83 ± 2.99
Disallowed	1.42 ± 1.93	1.82 ± 2.19	1.50 ± 2.12	1.77 ± 2.26	1.97 ± 2.29
Structure *Z*-score distribution[b]					
1st-generation packing quality	−3.48 ± 1.71	−2.41 ± 1.73	**−2.14 ± 1.72**	−2.37 ± 1.97	−2.44 ± 2.03
2nd-generation packing quality	−2.96 ± 1.19	**−1.28 ± 1.50**	−1.89 ± 1.12	−2.09 ± 1.27	−2.08 ± 1.26
Ramachandran Plot Appearance	−5.85 ± 1.92	**0.98 ± 2.03**	−4.13 ± 1.44	−4.31 ± 1.60	−4.43 ± 1.59
χ_1–χ_2 rotamer normality	−6.35 ± 2.35	**−1.68 ± 1.68**	−2.58 ± 1.59	−2.50 ± 1.42	−2.77 ± 1.40
Backbone conformation	−1.67 ± 1.50	**−1.07 ± 1.42**	−1.50 ± 1.40	−1.64 ± 1.37	−1.63 ± 1.33
RMS *Z*-score distribution[c]	0.83 ± 0.34	0.85 ± 0.06	0.84 ± 0.14	0.84 ± 0.13	**0.86 ± 0.13**
Bond lengths	1.06 ± 0.42	**1.03 ± 0.12**	0.78 ± 0.12	0.78 ± 0.11	0.80 ± 0.11
Bond angles	0.16 ± 0.27	1.31 ± 0.37	0.75 ± 0.15	0.74 ± 0.15	**0.75 ± 0.14**
Ω Angle restraints	1.71 ± 1.42	0.70 ± 0.37	**0.81 ± 0.18**	1.27 ± 0.54	1.31 ± 0.58
Side-chain planarity	0.92 ± 0.39	0.79 ± 0.17	**1.01 ± 0.16**	1.19 ± 0.36	1.23 ± 0.38
Improper dihedral distribution	1.07 ± 0.13	**1.06 ± 0.12**	1.07 ± 0.12	1.08 ± 0.13	1.08 ± 0.13

[a]Bold font indicates the best scores.
[b]Positive is better than average.
[c]RMS *Z*-score should be close to 1.



			Original	Refinement	DRESS	RECOORD CNW	RECOORD CYW
Steric Clash Score			228 ± 0	0 ± 0	45.28 ± 0	28.94 ± 3.3	31.4 ± 4.01
RMS Violation of All distance restraints			0.18 ± 0	0.13 ± 0	0 ± 0	0 ± 0	0 ± 0
Optimal Protein Energy	DOPE Score		-5956.02 ± 0	-6758.34 ± 0	-6996.23 ± 0	-6885.12 ± 80.48	-6827.91 ± 91.72
	Normalized DOPE Score		-0.46 ± 0	-1.28 ± 0	-1.53 ± 0	-1.42 ± 0.08	-1.36 ± 0.09
	dDFIRE Score		-119.83 ± 0	-136.4 ± 0	-134.44 ± 0	-133.38 ± 1.68	-133.49 ± 1.69
Ramachandran Indicators Ratio	Molprobity	Favorable	70.15 ± 0	97.01 ± 0	92.54 ± 0	89.91 ± 2.8	88 ± 2.57
		Allowed	22.39 ± 0	1.49 ± 0	5.97 ± 0	8.78 ± 2.52	10.09 ± 2.49
		Disallowed	7.46 ± 0	1.49 ± 0	1.49 ± 0	1.31 ± 0.79	1.91 ± 0.92
	PROCHECK	Favorable	62.9 ± 0	88.7 ± 0	80.6 ± 0	78.05 ± 2.59	76.57 ± 2.66
		Allowed	32.3 ± 0	9.7 ± 0	17.7 ± 0	20.14 ± 2.96	20.99 ± 2.87
		Generously	4.8 ± 0	0 ± 0	1.6 ± 0	1.79 ± 1.25	2.37 ± 1.75
		Disallowed	0 ± 0	1.6 ± 0	0 ± 0	0 ± 0	0.06 ± 0.32
Structure Z-Scores Distribution*	1st Generation Packing Quality		-1.18 ± 0	-1.07 ± 0	0.55 ± 0	-0.05 ± 0.45	-0.08 ± 0.37
	2nd Generation Packing Quality		-2.49 ± 0	-1.58 ± 0	-1.04 ± 0	-1.48 ± 0.44	-1.45 ± 0.39
	Ramachandran Plot Appearance		-7.28 ± 0	0.01 ± 0	-4.38 ± 0	-4.51 ± 0.6	-4.62 ± 0.51
	χ1-χ2 chi1_chi2 Normality		-8.71 ± 0	-4.88 ± 0	-4.78 ± 0	-5.44 ± 0.67	-5.53 ± 0.53
	Backbone Conformation		-0.94 ± 0	-1.76 ± 0	0.49 ± 0	0.14 ± 0.48	0 ± 0.61
RMS Z-Scores Distribution	Bond Lengths		0.96 ± 0	0.83 ± 0	1.15 ± 0	1.08 ± 0.03	1.11 ± 0.04
	Bond Angles		0.97 ± 0	1.18 ± 0	0.95 ± 0	0.9 ± 0.03	0.95 ± 0.03
	Ω Angle Restraints		0.15 ± 0	1.68 ± 0	1.06 ± 0	0.98 ± 0.06	1 ± 0.05
	Side Chain Planarity		2.62 ± 0	0.69 ± 0	0.91 ± 0	1.18 ± 0.28	1.15 ± 0.33
	Improper dihedral Distribution		0.88 ± 0	1.03 ± 0	1.33 ± 0	1.25 ± 0.07	1.3 ± 0.08
	Indide / Outside Distribution		1.01 ± 0	1 ± 0	1 ± 0	0.99 ± 0.01	0.99 ± 0.02
Individual ensemble			Link	Link			

Figure 2. General information on the target NMR structure: Jmol viewer, link-out to other databases, comparison of original and refinement structures.

will be disabled), and the NCBI Taxonomy Browser. (iii) The Results page shows the calculated results for the structural quality assessment based on the original and refined structures. It sometimes provides information on the evaluation results with DRESS and RECOORD with a link to the target protein sites and an overall summary table for the target structure. (iv) Visualization information on the details of the target structure and Ramachandran plot for protein quality assessment are provided. Both the original and refined structures are displayed (detailed view of Ramachandran indicators, location of steric clash sites, etc.), and the location of the Ramachandran indicator is shown on the plot with KiNG display software (27). (v) A download area is available for various data, for example, the result of the protein quality assessment for the original and refined structures, both PDB files, and the visualization inputs for the KiNG display.

ACKNOWLEDGEMENTS

PLSI supercomputing resources of Korea Institute of Sciences and Technology Information supported this work.

FUNDING

Funding for open access charge: A grant from the 'Korean Research Institute of Bioscience and Biotechnology Research Initiative Program'; and the Korean Ministry of Education, Science and Technology (MEST) under grant numbers 20110002321 and 20110019747, respectively.

Conflict of interest statement. None declared.

REFERENCES

1. Rose,P.W., Beran,B., Bi,C., Bluhm,W.F., Dimitropoulos,D., Goodsell,D.S., Prlic,A., Quesada,M., Quinn,G.B., Westbrook,J.D. *et al*. (2011) The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic Acids Res*., **39**, D392–D401.
2. Penkett,C.J., van Ginkel,G., Velankar,S., Swaminathan,J., Ulrich,E.L., Mading,S., Stevens,T.J., Fogh,R.H., Gutmanas,A., Kleywegt,G.J. *et al*. (2010) Straightforward and complete deposition of NMR data to the PDBe. *J. Biomol. NMR*, **48**, 85–92.
3. Melnik,B.S., Garbuzinskii,S.A., Lobanov,M. and Galzitskaia,O.V. (2005) The difference between protein structures that are obtained by X-ray analysis and magnetic resonance spectroscopy. *Mol. Biol*., **39**, 129–138.
4. Spronk,C.A., Linge,J.P., Hilbers,C.W. and Vuister,G.W. (2002) Improving the quality of protein structures derived by NMR spectroscopy. *J. Biomol. NMR*, **22**, 281–289.
5. Clore,G.M. and Gronenborn,A.M. (1998) New methods of structure refinement for macromolecular structure determination by NMR. *Proc. Natl Acad. Sci. USA*, **95**, 5891–5898.
6. Chen,J., Im,W. and Brooks,C.L. III (2004) Refinement of NMR structures using implicit solvent and advanced sampling techniques. *J. Am. Chem. Soc*., **126**, 16038–16047.
7. Linge,J.P., Williams,M.A., Spronk,C.A., Bonvin,A.M. and Nilges,M. (2003) Refinement of protein structures in explicit solvent. *Proteins*, **50**, 496–506.
8. Laskowski,R.A., Rullmannn,J.A., MacArthur,M.W., Kaptein,R. and Thornton,J.M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.
9. Hooft,R.W., Vriend,G., Sander,C. and Abola,E.E. (1996) Errors in protein structures. *Nature*, **381**, 272.
10. Davis,I.W., Leaver-Fay,A., Chen,V.B., Block,J.N., Kapral,G.J., Wang,X., Murray,L.W., Arendall,W.B. III, Snoeyink,J., Richardson,J.S. *et al*. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res*., **35**, W375–W383.
11. Nabuurs,S.B., Nederveen,A.J., Vranken,W., Doreleijers,J.F., Bonvin,A.M., Vuister,G.W., Vriend,G. and Spronk,C.A. (2004) DRESS: a database of REfined solution NMR structures. *Proteins*, **55**, 483–486.
12. Nederveen,A.J., Doreleijers,J.F., Vranken,W., Miller,Z., Spronk,C.A., Nabuurs,S.B., Guntert,P., Livny,M., Markley,J.L., Nilges,M. *et al*. (2005) RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins*, **59**, 662–672.
13. Kuszewski,J., Gronenborn,A.M. and Clore,G.M. (1996) Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases. *Protein Sci*., **5**, 1067–1080.
14. Stein,E.G., Rice,L.M. and Brunger,A.T. (1997) Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. *J. Magn. Reson*., **124**, 154–164.
15. Sippl,M.J. (1990) Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J. Mol. Biol*., **213**, 859–883.
16. Ulrich,E.L., Akutsu,H., Doreleijers,J.F., Harano,Y., Ioannidis,Y.E., Lin,J., Livny,M., Mading,S., Maziuk,D., Miller,Z. *et al*. (2008) BioMagResBank. *Nucleic Acids Res*., **36**, D402–D408.
17. Lazaridis,T. and Karplus,M. (1999) Effective energy function for proteins in solution. *Proteins*, **35**, 133–152.
18. Brooks,B.R., Brooks,C.L. III, Mackerell,A.D. Jr, Nilsson,L., Petrella,R.J., Roux,B., Won,Y., Archontis,G., Bartels,C., Boresch,S. *et al*. (2009) CHARMM: the biomolecular simulation program. *J. Comput. Chem*., **30**, 1545–1614.
19. Shen,M.Y. and Sali,A. (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci*., **15**, 2507–2524.
20. Chen,H. and Kihara,D. (2008) Estimating quality of template-based protein models by alignment stability. *Proteins*, **71**, 1255–1274.
21. Yang,Y. and Zhou,Y. (2008) Specific interactions for ab initio folding of protein terminal regions with secondary structures. *Proteins*, **72**, 793–803.
22. Doreleijers,J.F., Raves,M.L., Rullmann,T. and Kaptein,R. (1999) Completeness of NOEs in protein structure: a statistical analysis of NMR. *J. Biomol. NMR*, **14**, 123–132.
23. Laskowski,R.A. (2003) Structural quality assurance. *Methods Biochem. Anal*., **44**, 273–303.
24. Shaanan,B., Gronenborn,A.M., Cohen,G.H., Gilliland,G.L., Veerapandian,B., Davies,D.R. and Clore,G.M. (1992) Combining experimental information from crystal and solution studies: joint X-ray and NMR refinement. *Science*, **257**, 961–964.
25. Wichmann,B.A. and Hill,I.D. (1982) Algorithms AS 183: an efficient and portable pseudo-random number generator. *Appl. Statist*., **31**, 188–190.
26. Herraez,A. (2006) Biomolecules in the computer: Jmol to the rescue. *Biochem. Mol. Biol. Educ*., **34**, 255–261.
27. Chen,V.B., Davis,I.W. and Richardson,D.C. (2009) KING (Kinemage, Next Generation): a versatile interactive molecular and scientific visualization program. *Protein Sci*., **18**, 2403–2409.